JOURNAL OF MULTIVARIATE ANALYSIS 48, 107–114 (1994)

An Identity for the Noncentral Wishart Distribution with Application

PUI LAM LEUNG

The Chinese University of Hong Kong, Shatin, Hong Kong

This paper generalizes an identity for the Wishart distribution (derived independently by C. Stein and L. Haff) to the noncentral Wishart distribution. As an application of this noncentral Wishart identity, we consider the problem of estimating the noncentrality matrix of a noncentral Wishart distribution. This noncentral Wishart identity is used to develop a class of orthogonally invariant estimators which dominate the usual unbiased estimator. © 1994 Academic Press, Inc.

1. INTRODUCTION

Let S be an $m \times m$ positive definite random matrix having a Wishart distribution with n degrees of freedom and covariance matrix Σ, denoted by $S = (s_{ij}) \sim W_m(n, \Sigma)$. A scalar identity concerning the expectation of functions of S was given in Haff [6]. For reference, we define the following symbols and state this Wishart identity.

Let V be a matrix and $V_{(r)} = rV + (1-r)\operatorname{diag}(V)$. We define

$$D = (d_{ij}) = \tfrac{1}{2}(1 + \delta_{ij})\, \partial/\partial s_{ij} \tag{1.1}$$

as a matrix of differential operators, where δ_{ij} is the Kronecker delta. DV is the formal matrix product of D and V and $\partial h(S)/\partial S = (\partial h(S)/\partial s_{ij})$ for a real-valued function $h(S)$. Under fairly general regularity conditions, we have

$$\begin{aligned} E[h(S)\operatorname{tr}(\Sigma^{-1}V)] = {} & 2E[h(S)\operatorname{tr}(DV)] + 2E\operatorname{tr}[(\partial h(S)/\partial S)\, V_{(1/2)}] \\ & + (n-m-1)\, E[h(S)\operatorname{tr}(S^{-1}V)], \end{aligned} \tag{1.2}$$

where V is a matrix whose elements are functions of S and Σ, $h(S)$ is a real-valued function of S. This Wishart identity is derived from Stokes' theorem, a multivariate integration by parts. The regularity conditions are

Received September 1, 1992; revised March 26, 1993.

Key words and phrases: Wishart identity, noncentral Wishart identity, noncentrality matrix, decision-theoretic estimation.

0047-259X/94 $6.00

to ensure that the function hV satisfies the conditions of Stokes' theorem and are given in Haff [2].

Haff used this identity to prove dominance results in a decision-theoretic estimation problem in his papers [3, 4]. Later on, Haff [5] extended (1.2) to matrix identities with applications in regression as well. The Wishart identity was also used in Leung and Muirhead [10]. Muirhead and Verathaworn [13] obtained an identity for the multivariate F distribution similar to (1.2). Applications of this F identity can also be found in Leung [9], Konno [7, 8].

In this paper, we generalize (1.2) to the noncentral Wishart distribution, and the result is called the "noncentral Wishart identity." In Section 2, the noncentral Wishart identity is stated and proved. As an application of this identity, we consider the problem of estimating the noncentrality matrix Δ of a noncentral Wishart distribution in Section 3. A class of orthogonally invariant estimators of Δ which dominates the usual unbiased estimator of Δ is proposed. This problem was also considered by Leung and Muirhead [10].

2. The Noncentral Wishart Identity

Assume that A is an $m \times m$ positive definite random matrix having a noncentral Wishart distribution with n degrees of freedom, covariance matrix Σ, and noncentrality matrix Δ, denoted by $A \sim W_m(n, \Sigma, \Delta)$. Let $V(A, \Delta)$ be a matrix whose elements are functions of A and Δ, and $h(A)$ be a real-valued function of A. We simply write $V(A, \Delta)$ as V and $h(A)$ as h for brevity. Under the same regularity conditions for the Wishart identity given in Haff [2], we have

Theorem 2.1 (Noncentral Wishart identity).

$$\begin{aligned} E[h \operatorname{tr}(\Sigma^{-1} V)] = 2E[h \operatorname{tr}(DV)] &+ 2E \operatorname{tr}[(\partial h(A)/\partial A)\, V_{(1/2)}] \\ &+ (n-m-1)\, E[h \operatorname{tr}(A^{-1} V)] \\ &+ E_1[h \operatorname{tr}(\Delta A^{-1} V)], \end{aligned} \tag{2.1}$$

where the expectation E is taken over a $W_m(n, \Sigma, \Delta)$ distribution and the expectation E_1 in the last term of (2.1) is taken over a $W_m(n+m+1, \Sigma, \Delta)$ distribution, i.e.,

$$E_1[h \operatorname{tr}(\Delta A^{-1} V)] = \int_{A>0} h \operatorname{tr}(\Delta A^{-1} V) f_1(A)(dA)$$

with $f_1(A)$ denoting the density of a $W_m(n+m+1, \Sigma, \Delta)$ distribution. The identity (2.1) is the same as (1.2) except for the last term, where the expectation is taken over a noncentral Wishart distribution with degrees of freedom changed from n to $n+m+1$. Note that when $\Delta = 0$, (2.1) reduces to (1.2).

Before we prove (2.1), we need the following Lemma.

LEMMA 2.2. (i) $E[\operatorname{tr}(\Sigma^{-1}A)] = nm + \operatorname{tr}\Delta$,

(ii) $E[\operatorname{tr}(\Sigma^{-1}A\Sigma^{-1}A)] = \operatorname{tr}(\Delta^2) + 2(n+m+1)(\operatorname{tr}\Delta) + nm(n+m+1)$,

(iii) $\operatorname{tr} D(A\Sigma^{-1}A) = (m+1)\operatorname{tr}(\Sigma^{-1}A)$.

Proof. Part (i) can be easily obtained by noting that $\Sigma^{-1/2}A\Sigma^{-1/2}$ is distributed as $W(n, I, \Sigma^{1/2}\Delta\Sigma^{-1/2})$ and (ii) follows from Theorem 4.4 of Magnus and Neudecker [11]. Part (iii) is given in Lemma (2.2) of Konno [7].

Now we turn to the proof of Theorem 2.1.

Proof of (2.1). The proof is similar to the proof of (1.2) given in Haff [2]. The density of A is

$$f(A) = C(\det A)^{(n-m-1)/2}\, e\operatorname{tr}(-\Sigma^{-1}A/2)\, {}_0F_1(n/2; \Delta\Sigma^{-1}A/4), \qquad (2.2)$$

where $C = [2^{mn/2}\Gamma_m(n/2)]^{-1}(\det\Sigma)^{-n/2}\, e\operatorname{tr}(-\Delta/2)$, $n-m-1>0$, $A>0$, $e\operatorname{tr}(\cdot) = \exp[\operatorname{tr}(\cdot)]$, ${}_0F_1(\cdot)$ is the hypergeometric function with matrix argument, and $\Gamma_m(\cdot)$ is the multivariate gamma function (see Muirhead [12] for details). For the differential operator D defined in (1.1), we have

$$D(\det A)^{(n-m-1)/2} = [(n-m-1)/2](\det A)^{(n-m-1)/2} A^{-1}$$

and

$$De\operatorname{tr}(-\Sigma^{-1}A/2) = (-\Sigma^{-1}/2)\, e\operatorname{tr}(-\Sigma^{-1}A/2).$$

Hence, D operating on $f(A)$ in (2.2) gives

$$Df(A) = \left[\frac{n-m-1}{2}A^{-1} - \frac{1}{2}\Sigma^{-1}\right] f(A) + Q(A), \qquad (2.3)$$

where

$$Q(A) = C(\det A)^{(n-m-1)/2}\, e\operatorname{tr}(-\Sigma^{-1}A/2)[D\, {}_0F_1(n/2; \Delta\Sigma^{-1}A/4)]. \qquad (2.4)$$

The same set of regularity conditions on hV given in Haff [2] ensures that $\int_{A>0} \operatorname{tr} D[hVf(A)](dA) = 0$. It follows that

$$0 = E\operatorname{tr}[(\partial h/\partial A)\, V_{(1/2)}] + E[h\operatorname{tr}(DV)] + \int_{A>0} h\operatorname{tr}[V(Df)](dA).$$

Using (2.3), we have

$$E[h\,\mathrm{tr}(\Sigma^{-1}V)] = 2E[h\,\mathrm{tr}(DV)] + 2E\,\mathrm{tr}[(\partial h(A)/\partial A)\,V_{(1/2)}] + (n-m-1)\,E[h\,\mathrm{tr}(A^{-1}V)] + 2\int_{A>0} h\,\mathrm{tr}(QV)(dA). \tag{2.5}$$

Comparing (2.5) with (2.1), we see that the proof is complete if

$$2\int_{A>0} h\,\mathrm{tr}(QV)(dA) = \int_{A>0} h\,\mathrm{tr}(\Delta A^{-1}V) f_1(A)(dA),$$

where $f_1(A)$ is the density of a $W_m(n+m+1, \Sigma, \Delta)$ distribution. Therefore, it suffices to show that

$$2Q(A) f_1^{-1}(A) = \Delta A^{-1} \qquad \text{a.e.} \tag{2.6}$$

$Q(A)$ defined in (2.4) involves the operation of D on ${}_0F_1(\cdot)$. Although it is possible to prove (2.6) directly by differentiating the series of zonal polynomials in ${}_0F_1(\cdot)$, this could be very complicated and messy. We take another approach to proving (2.6). By taking $V(A, \Sigma) = A\Sigma^{-1}A$ and $h(A) = 1$ in (2.5) and from Lemma 2.2, we have

$$\begin{aligned}\mathrm{tr}(\Delta^2) + 2(n+m+1)(\mathrm{tr}\,\Delta) + nm(n+m+1) &= 2(m+1)(nm + \mathrm{tr}\,\Delta) + (n-m-1)(nm + \mathrm{tr}\,\Delta) \\ &\quad + 2\int_{A>0} \mathrm{tr}(QA\Sigma^{-1}A)(dA),\end{aligned}$$

which implies

$$2\int_{A>0} \mathrm{tr}(QA\Sigma^{-1}A)(dA) = \mathrm{tr}(\Delta^2) + (n+m+1)(\mathrm{tr}\,\Delta). \tag{2.7}$$

Note that the right-hand side of (2.7) is equal to

$$\int_{A>0} \mathrm{tr}(\Delta\Sigma^{-1}A) f_1(A)(dA),$$

where $f_1(A)$ is the density of a $W_m(n+m+1, \Sigma, \Delta)$ distribution.

It follows from (2.7) that $\mathrm{tr}[2QA\Sigma^{-1}Af_1^{-1}(A) - \Delta\Sigma^{-1}A] = 0$ a.e. or $\mathrm{tr}\{[2Qf_1^{-1}(A) - \Delta A^{-1}](A\Sigma^{-1}A)\} = 0$ a.e. for all $A > 0$ and $\Sigma > 0$, which implies (2.6), and the proof is completed.

Remark. It is possible to obtain a similar identity for a noncentral multivariate F distribution but this is not explored in the present paper.

3. Improved Estimation of Noncentrality Matrix

The Wishart identity (1.2) is very useful for finding bounds for expectations which often appear in risk calculations. We expect that similar applications can be found for the noncentral Wishart identity (2.1) as well. To illustrate a nontrivial application of the noncentral Wishart identity, we consider the problem of estimating the noncentrality matrix of a noncentral Wishart distribution. Assume that $A \sim W_m(n, I, \Delta)$, i.e., A has a noncentral Wishart distribution with n degrees of freedom, identity covariance matrix, and noncentrality matrix Δ. The usual unbiased estimator of Δ is $\hat{\Delta}_{\mathrm{U}} = A - nI$. By using a decision-theoretic approach, we estimate Δ using the invariant loss function

$$L(\Delta, \hat{\Delta}) = \operatorname{tr}(\Delta^{-1}\hat{\Delta} - I_m)^2. \tag{3.1}$$

Let $\hat{\Delta}_\alpha = \hat{\Delta}_{\mathrm{U}} + (\alpha/(\operatorname{tr} A))\, I_m$. It is shown in Theorem 3.2 that $\hat{\Delta}_\alpha$ dominate $\hat{\Delta}_{\mathrm{U}}$ for a suitable choice of α. This problem and $\hat{\Delta}_\alpha$ were also considered by Leung and Muirhead [10], using the squared error loss function $L(\Delta, \hat{\Delta}) = \operatorname{tr}(\hat{\Delta} - \Delta)^2$ (see Leung and Muirhead [10] for the significance and details of this problem). The univariate version of this problem was considered by Perlman and Rasmussen [14], Saxena and Alam [15], and Chow [1]. Before we state and prove the dominance result, we need the following lemma.

Lemma 3.1. *Assume that $n > 4$. Then*

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{\operatorname{tr} A}\right] \leqslant nE\left[\frac{\operatorname{tr} \Delta^{-2}}{\operatorname{tr} A}\right] - 2(n-4)\, E\left[\frac{\operatorname{tr} \Delta^{-2}}{(\operatorname{tr} A)^2}\right] + E_1\left[\frac{\operatorname{tr} \Delta^{-1}}{\operatorname{tr} A}\right] - 2E_1\left[\frac{\operatorname{tr} \Delta^{-1}}{(\operatorname{tr} A)^2}\right],$$

where E is taken over a $W_m(n, I, \Delta)$ distribution and E_1 is taken over a $W_m(n+m+1, I, \Delta)$ distribution.

Proof. We apply the noncentral Wishart identity given in (2.1) with $\Sigma = I$, $V = \Delta^{-2}A$, and $h = 1/(\operatorname{tr} A)$. Since $\operatorname{tr}(DV) = [(m+1)/2](\operatorname{tr} \Delta^{-2})$ and $\partial h/\partial A = [-1/(\operatorname{tr} A)^2]\, I_m$ (see Haff [4]), we have

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{\operatorname{tr} A}\right] = nE\left[\frac{\operatorname{tr} \Delta^{-2}}{\operatorname{tr} A}\right] - 2E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{(\operatorname{tr} A)^2}\right] + E_1\left[\frac{\operatorname{tr} \Delta^{-1}}{\operatorname{tr} A}\right]. \tag{3.2}$$

To compute the second term on the right-hand side of (3.2), we apply identity (2.1) again with $V=\Delta^{-2}A$ and $h=1/(\operatorname{tr} A)^2$. Since $\partial h/\partial A=[-2/(\operatorname{tr} A)^3]\, I_m$ (see Haff [4]), (3.2) becomes

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{(\operatorname{tr} A)^2}\right]=nE\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} A)^2}\right]-4E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{(\operatorname{tr} A)^3}\right]+E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} A)^2}\right]. \quad (3.3)$$

Using the fact that $\operatorname{tr}(\Delta^{-2}A)\leqslant(\operatorname{tr}\Delta^{-2})(\operatorname{tr} A)$ in the second term of the right-hand side of (3.3), we have

$$E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{(\operatorname{tr} A)^2}\right]\geqslant(n-4)\,E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} A)^2}\right]+E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} A)^2}\right]. \quad (3.4)$$

With (3.4) substituted into (3.2), the proof is completed.

THEOREM 3.2. *Assume that* $n>4$. *Then* $\hat{\Delta}_\alpha$ *dominate* $\hat{\Delta}_{\mathrm{U}}$ *if* $0<\alpha<4(n-4)$.

Proof. For the loss function defined in (3.1), it is straightforward to show that the risk difference between $\hat{\Delta}_{\mathrm{U}}$ and $\hat{\Delta}_\alpha$ is

$$\begin{aligned}G(\Delta)&=E[L(\Delta,\hat{\Delta}_{\mathrm{U}})-L(\Delta,\hat{\Delta}_\alpha)]\\&=2\alpha E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} A}\right]-2\alpha E\left[\frac{\operatorname{tr}(\Delta^{-2}A)}{\operatorname{tr} A}\right]+2n\alpha E\left[\frac{\operatorname{tr}\Delta^{-2}}{\operatorname{tr} A}\right]-\alpha^2E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} A)^2}\right].\end{aligned}$$

Using Lemma 3.1 and simplifying, we have

$$\begin{aligned}G(\Delta)\geqslant{}&2\alpha E\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} A}\right]-2\alpha E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{\operatorname{tr} A}\right]+4\alpha E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} A)^2}\right]\\&+\alpha[4(n-4)-\alpha]\,E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} A)^2}\right]. \quad (3.5)\end{aligned}$$

From Eq. (2.9) in Leung and Muirhead [9], $E[1/(\operatorname{tr} A)]=E_2[1/(mn+2K-2)]$ and $E_1[1/(\operatorname{tr} A)]=E_2\{1/[m(n+m+1)+2K-2]\}$, where K has a Poisson distribution with mean $\frac{1}{2}(\operatorname{tr}\Delta)$ and E_2 is taken with respect to K. Therefore (3.5) becomes

$$\begin{aligned}G(\Delta)\geqslant{}&2\alpha(\operatorname{tr}\Delta^{-1})\,E_2\left[\frac{m(m+1)}{(mn+2K-2)[m(n+m+1)+2K-2]}\right]\\&+4\alpha E_1\left[\frac{\operatorname{tr}\Delta^{-1}}{(\operatorname{tr} A)^2}\right]+\alpha[4(n-4)-\alpha]\,E\left[\frac{\operatorname{tr}\Delta^{-2}}{(\operatorname{tr} A)^2}\right].\end{aligned}$$

The first two terms on the right-hand side are nonnegative. A sufficient condition for $G(\Delta)\geqslant 0$ is $0\leqslant\alpha\leqslant 4(n-4)$ and the proof is completed.

Remark. $\hat{\Delta}_U$ and $\hat{\Delta}_\alpha$ are not necessarily positive definite. They are dominated by their truncated versions $\hat{\Delta}_U^+$ and $\hat{\Delta}_\alpha^+$, respectively; i.e., $\hat{\Delta}^+$ is formed by replacing the negative eigenvalues in $\hat{\Delta}$ by zero. The following theorem demonstrates this.

THEOREM 3.3. *Let $\hat{\Delta}$ be any estimator Δ and $\hat{\Delta}^+$ be the truncated version of $\hat{\Delta}$. Then $E[L(\Delta, \hat{\Delta})] \geqslant E[L(\Delta, \hat{\Delta}^+)]$ for all Δ and the inequality is strict if $\hat{\Delta}$ is not nonnegative definite.*

Proof. Let $\hat{\Delta} = HLH'$, where H is orthogonal and L is a diagonal matrix with diagonal elements $l_1, ..., l_m$. Then $\hat{\Delta}^+ = HL^+H'$, where $L^+ = \operatorname{diag}(l_1^+, ..., l_m^+)$ with $l_i^+ = l_i$ if $l_i > 0$ and $l_i^+ = 0$ if $l_i \leqslant 0$.

$$\begin{aligned} E[L(\Delta, \hat{\Delta})] &= E[\operatorname{tr}(\Delta^{-1}\hat{\Delta} - I_m)^2] \\ &= E \operatorname{tr}[\Delta^{-1/2}(\hat{\Delta} - \Delta)\, \Delta^{-1/2}]^2 \\ &= E \operatorname{tr}[\Delta^{-1/2}H(L - H'\Delta H)\, H'\Delta^{-1/2}]^2 \\ &\geqslant E \operatorname{tr}[\Delta^{-1/2}H(L^+ - H'\Delta H)\, H'\Delta^{-1/2}]^2 \\ &= E[L(\Delta, \hat{\Delta}^+)]. \end{aligned}$$

The inequality is strict if $l_i < 0$ for some i, and the proof is completed.

REFERENCES

[1] CHOW, M. S. (1987). A complete class theorem for estimating a noncentrality parameter. *Ann. Statist.* **15** 800–804.

[2] HAFF, L. R. (1979). An identity for the Wishart distribution with applications. *J. Multivariate Anal.* **9** 531–544.

[3] HAFF, L. R. (1979). Estimation of the inverse covariance matrix: Random mixtures of the inverse Wishart matrix and the identity. *Ann. Statist.* **7** 1264–1276.

[4] HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *Ann. Statist.* **8** 586–597.

[5] HAFF, L. R. (1981). Further identities for the Wishart distribution with applications in regression. *Canad. J. Statist.* **9** 215–224.

[6] HAFF, L. R. (1982). Identities for the inverse Wishart distribution with computational results in linear and quadratic discrimination. *Sankhyā* **44** 245–258.

[7] KONNO, Y. (1988). Exact moments of the multivariate F and beta distributions. *J. Japan Statist. Soc.* **18** 123–130.

[8] KONNO, Y. (1991). A note on estimating eigenvalues of scale matrix of the multivariate F-distribution. *Ann. Inst. Statist. Math.* **43** 157–165.

[9] LEUNG, P. L. (1992). Estimation of the scale matrix of the multivariate F distribution. *Comm. Statist. Theory Methods* **21** 1845–1856.

[10] LEUNG, P. L., AND MUIRHEAD, R. J. (1987). Estimation of parameter matrices and eigenvalues in MANOVA and canonical correlation analysis. *Ann. Statist.* **15** 1651–1666.

[11] MAGNUS, J. R., AND NEUDECKER, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* 7 381–394.

[12] MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

[13] MUIRHEAD, R. J., AND VERATHAWORN, T. (1985). On estimating the latent roots of $\Sigma_1 \Sigma_2^{-1}$. In *Multivariate Analysis VI* (P. R. Krishnaiah, Ed.), pp. 431–447. North-Holland, Amsterdam.

[14] PERLMAN, M. D., AND RASMUSSEN, U. A. (1975). Some remarks on estimating a noncentrality parameter. *Comm. Statist.* **4** 455–468.

[15] SAXENA, K. M. L., AND ALAM, K. (1982). Estimation of the noncentrality parameter of a chi-squared distribution. *Ann. Statist.* **10** 1012–1016.